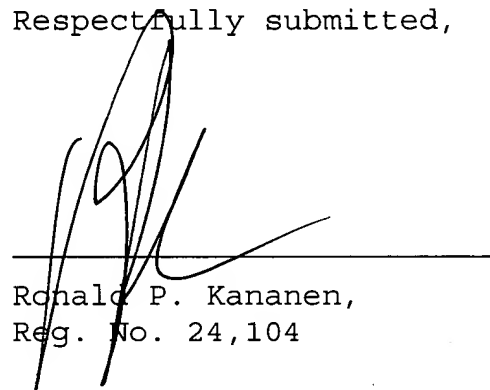


REMARKS

This Preliminary Amendment is requested prior to the initial examination of the above-identified patent application to address minor matters of form and syntax. No new matter has been added. If the Examiner has any suggestions for placing this application in even better form, the Examiner is invited to telephone the undersigned at the number listed below.

Respectfully submitted,

Date: July 16, 2003



Ronald P. Kananen,
Reg. No. 24,104

RADER, FISHMAN & GRAUER, PLLC
THE LION BUILDING
1233 20TH STREET, N.W., SUITE 501
WASHINGTON, D.C. 20036
TEL: (202) 955-3750
FAX: (202) 955-3751
CUSTOMER No. 23353



80001-2107
REDLINE VERSION

Attorney Docket No.: SON-2107
Application No.: 09/862,421

TITLE OF THE INVENTION

Picture Information Conversion Method and Apparatus

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to a method and apparatus for converting ~~the~~ picture information. More particularly, it relates to a picture information conversion method and apparatus for use in receiving ~~the~~ picture information of, e.g., MPEG pictures compressed by orthogonal transform, such as discrete cosine transform, and motion compression (bitstream), over a broadcast satellite, cable TV or a network medium, such as ~~Internet~~ the Internet, or in processing the bitstream on a recording medium, such as an optical disc or a magneto-optical disc.

Description of Related Art

There has so far been presented a picture information compression system, such as MPEG, for compressing the picture information by motion compression, by exploiting the redundancy proper to the picture information, and with a view to handling the picture information as digital data and to high-efficiency transmission and storage of the information. The apparatus conforming to this picture information compression method is finding widespread use in information distribution by, e.g., a broadcasting station and in information reception in homes.

In particular, the MPEG2 (ISO/IEC 13818-2) is defined as being a

comprehensive picture encoding system applicable to both ~~the~~ interlaced and progressive scanned pictures, to standard definition pictures, and to high-definition pictures.

That is, in the MPEG2 encoding compression system, codes of a bitrate of 4 to 8 Mbps are allocated to an interlaced scanned picture of a standard resolution with 720×480 pixels, and codes of a bitrate of 18 to 22 Mbps are allocated to a progressive scanned picture of a standard resolution with 1920×1088 pixels to realize a high compression factor and a high picture quality.

In light of the above, the MPEG2 is estimated to continue to be in extensive use in professional and consumer use.

However, the MPEG2 is mainly intended for high picture quality encoding for broadcasting, while it is not adapted to a coderate lower than that in MPEG1, that is, it is not adapted to an encoding system with a higher compression rate.

On the other hand, it may be predicted that the needs for an encoding system with a higher compression rate will continue to ~~be increased~~ increase ~~at a high compression rate~~. In order to cope with this situation, standardization of the MPEG4 encoding system with a high compression rate is underway. For this picture encoding system, the international standardization was acknowledged in December 1998 as ISO/IEC 14496-2.

Meanwhile, there also exists the needs for converting the MPEG2 compressed

picture information, once encoded for digital broadcast, into the compressed picture information (bitstream) of a lower code rate more amenable to processing on a portable terminal.

For accommodating these needs, there is presented a picture information converting apparatus (transcoder) in "Field-to-Frame Transcoding with Spatial and Temporal Downsampling" (Susie L. Wee, John G. Apostolopoulos, and Nick Feamster, ICIP 99; referred to below as reference 1).

As shown in Fig. 1, the picture information converting apparatus (transcoder) presented in this reference 1 is made up of a picture type decision unit 1, an MPEG2 picture information decoding unit (I/P picture) 2, a decimating unit 3, an MPEG2 picture information encoding unit (I/P-VOP) 4, a motion vector synthesis unit 5 and a motion vector detection unit 6.

This picture information converting apparatus is fed with the interlaced scanned MPEG2 compressed picture information (bitstream) made up of an intra-coded picture (I-picture) obtained on intra-frame coding, a forward predicted picture (P-picture) obtained on predictive coding by referring to a forward direction in the display sequence, and a bi-directionally coded picture (B-picture) obtained on predictive coding by referring to the forward and backward directions in the display sequence.

This MPEG2 compressed picture information (bitstream) is discriminated in the picture type decision unit 1 as to whether it is of an I/P picture or of a B-picture. Only

the I/P picture is output to the next following MPEG2 picture information decoding unit (I/P picture) 2, while the B-picture is discarded.

Similarly to the processing in a routine MPEG2 picture information decoding unit, the processing in the -MPEG2 picture information decoding unit (I/P picture) 2 decodes the MPEG2 compressed picture information (bitstream) into picture signals,

The pixel value output by the MPEG2 picture information decoding unit (I/P picture) 2 is input to the decimating unit 3, which then decimates the pixel value by 1/2 in the horizontal direction, while leaving only one of the data of the first field and the data of the second field ~~to~~ and discarding the other. By this decimation, there is produced a progressive scanned picture having a size equal to 1/4 of the input picture information.

The progressive scanned picture, generated by the decimating unit 3, is encoded by the MPEG2 picture information encoding unit (I/P-VOP) 4 into an intra-frame-coded I-VOP and ~~to~~ a P-VOP obtained on predictive coding by referring to the forward direction in the display sequence, and it is output as the MPEG4 compressed picture information (bitstream). Meanwhile, VOP means a video object plane and is equivalent to a frame in MPEG2.

The motion vector information in the input MPEG2 compressed picture information (bitstream) is mapped in the motion vector synthesis unit 5 into a motion

vector for the as-decimated picture information. The motion vector detection unit 6 detects the high precision motion vector based on the motion vector value synthesized in the motion vector synthesis unit 5.

The Reference 1 discusses on a picture information converting apparatus for generating the MPEG2 compressed picture information (bitstream) having a size equal to $1/2 \times 1/2$ of the input MPEG2 compressed picture information (bitstream). That is, if the input MPEG2 compressed picture information (bitstream) is in meeting with the NTSC (National Television System Committee), the output MPEG4 compressed picture information (bitstream) is of an SIF size (352×240 pixels).

Meanwhile, in the picture information converting apparatus, shown in Fig. 1, the code rate control in the MPEG4 picture information encoding unit (I/P-VOP) 4 represents a significant factor in determining the picture quality in the MPEG4 compressed picture information. In the ISO/IEC 14496-2, there is no particular definition as to the coderate controlling system, such that each vendor may use a system that is possibly optimal from the viewpoint of the processing volume and the output picture quality depending on the particular application. The system discussed in MPEG2 Test Model 5(ISO/IEC JTC1/SC29/WG11 NO400) is hereinafter explained as a typical coderate controlling system.

The code rate control flow is now explained by referring to the flowchart of Fig.2. At a first step S11, the picture information encoding unit (I/P-VOP) 4 allocates

bits to each picture, with the target code rate (target bitrate) and the GOP (group of pictures) as input variables. It is noted that a GOP means a set of pictures accessible at random.

That is, at step S11, the picture information encoding unit (I/P-VOP) 4 distributes bits to be allocated to each picture in the GOP, based on the volume of bits allocated to a picture not as yet decoded in the GOP inclusive of pictures intended for allocation. This bit volume is referred to below as R. This distribution is repeated in the sequence of the encoded pictures in the GOP. In this case, coderate allocation to each picture is made using two ~~suppositions~~ assumptions as now explained.

It is first assumed that the product of an average quantization scale code used in encoding each picture and the volume of the codes generated is unchanged from one picture type to another as long as the picture displayed is not changed. Based on this supposition, variables X_i , X_p and X_b representing the picture complexity (global complexity measure) are updated after encoding each picture in accordance with the following equation (1) from one picture type to another:

$$X_i = S_i \cdot Q_i$$

$$X_p = S_p \cdot Q_p$$

$$X_b = S_b \cdot Q_b$$

...(1)

It is noted that S_i , S_p and S_b denote the volumes of the codes generated on picture encoding, and Q_i , Q_p and Q_b are average quantization scale codes at the time of picture encoding. On the other hand, the initial value, in terms of the target bitrate bit_rate [bits/sec], is as indicated in the following equation (2):

$$X_i = 160 \times \text{bit_rate} / 115$$

$$X_p = 60 \times \text{bit_rate} / 115$$

$$X_b = 42 \times \text{bit_rate} / 115$$

...(2).

Second, it is assumed that the overall picture quality is optimized at all times when the proportions K_p , K_b of the quantization scale code of the P- and B-pictures, referenced to the quantization scale code of an I-picture, are of values defined in the following equation (3):

$$K_p = 1.0; K_b = 1.4$$

...(3).

That is, the quantization scale code of a B-picture is set at all times so as to be 1.4 times the quantization scale codes of the I- and P-pictures. This is, based on the supposition that if, the volume of the codes that can be saved in a B-picture by encoding the B-picture slightly more coarsely than the I- and P-pictures is added to the code volume of the I- and P-pictures, the I- and P-pictures can be improved in picture quality, so that the B-picture which refers to these also can also be improved in picture

From the above-mentioned two ~~suppositions~~ assumptions, the volumes of bits allocated to each picture of the GOP (T_i , T_p , T_b) are as indicated by the following equation (4):

$$T_i = \max \left\{ \frac{R}{1 + \frac{N_p \cdot X_p}{X_i \cdot K_p} + \frac{N_b \cdot X_b}{X_i \cdot K_b}}, \frac{bit_rate}{8 \times picture_rate} \right\}$$
$$T_p = \max \left\{ \frac{R}{N_p + \frac{N_b \cdot K_p \cdot X_b}{K_b \cdot X_p}}, \frac{bit_rate}{8 \times picture_rate} \right\}$$
$$T_b = \max \left\{ \frac{R}{N_b + \frac{N_p \cdot K_b \cdot X_p}{K_p \cdot X_b}}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

...(4).

where N_p and N_b denote the number of P- and B-pictures, respectively, not as yet encoded in the GOP.

Based on the value of the allocated codes, thus found, the volume of bits R allocated to uncoded pictures in a GOP is updated in accordance with the following equation (5): each time a picture is encoded in accordance with steps 511 and 512:

$$R = R - S_{i,p,b}$$

...(5)

~~each time a picture is encoded in accordance with steps S11 and S12.~~

On the other hand, in encoding the first picture of the GOP, R is updated in accordance with the equation (6):

$$R = \frac{\text{bit_rate} \times N}{\text{picture_rates}} + R$$

...(6)

where N is the number of pictures in a GOP. The initial value of R at the beginning of a sequence is 0.

Then, at step S12, the picture information encoding unit (I/P-VOP) 4 performs rate control using a virtual buffer. That is, at step S12, the picture information encoding unit (I/P-VOP) 4 finds the quantization scale code by macro-block based feedback control, based on the capacitance of three types of the virtual buffer as-set independently for the respective pictures, in order to make the volume of allocated bits for the respective pictures as-found by the equation (4) at step S11 (T_i , T_p , T_b) coincident with the actual volume of generated codes.

Before proceeding to the encoding of the jth macroblock, the occupancy volume of the virtual buffer is found by the following equation (7):

$$\begin{aligned} d_j^i &= d_o^i + B_{j-1} - \frac{T_i \times (j-1)}{MB_cnt} \\ d_j^p &= d_o^p + B_{j-1} - \frac{T_p \times (j-1)}{MB_cnt} \\ d_j^b &= d_o^b + B_{j-1} - \frac{T_b \times (j-1)}{MB_cnt} \end{aligned}$$

... (7)

It is noted that do^i , do^p , do^b are initial occupancy volumes of the virtual buffers, B_j is the volume of bits generated from the leading end of a picture up to the j th macroblock and MB_cnt is the number of macroblocks in one picture. The occupancy of the virtual buffer at the time of end of encoding of each picture (dMB_cnt^i , dMB_cnt^p , dMB_cnt^b) is used as initial values (do^i , do^p , do^b) of the occupancy of the virtual buffer for the next picture in the same picture type.

The quantization scale code for the j th macroblock is then calculated in accordance with the with equation (8):

$$Q_j = \frac{d_j \times 31}{r}$$

... (8)

where r is a variable controlling the feedback loop response, termed a reaction parameter, and is given by the equation (9):

$$r = 2 \times \frac{bit_rate}{picture_rate}$$

... (9).

Meanwhile, the initial value of the virtual buffer at the time of beginning of the encoding is given by the equation (10):

$$\begin{aligned}d_o^i &= 10 \times \frac{r}{31} \\d_o^p &= K_p \bullet d_o^i \\d_o^b &= K_{pb} \bullet d_o^i\end{aligned}\dots(10),$$

Finally, at S 13, the picture information encoding unit (I/P-VOP) 4 performs macro-block based adaptive quantization taking psychoacoustic characteristics into account. That is, at step S13, the picture information encoding unit (I/P-VOP) 4 varies the quantization scale code as found at step S12 by a variable termed macroblock-based activity in such a manner that the quantization scale code will be quantized finely and coarsely in a monotonous pattern portion where deterioration tends to be visually outstanding and in a complex pattern portion where deterioration is less likely to be outstanding, respectively.

The activity is given, using luminance signal pixel values of an original picture, four blocks in the frame DCT mode and four blocks in the field DCT mode, totaling at eight blocks, by the following equation (11):

$$\begin{aligned}act_j &= 1 + \min_{sblk=1,8} (var_sblk) \\var_sblk &= \frac{1}{64} \sum_{k=1}^{64} (P_k - \bar{P})^2 \\ \bar{P} &= \frac{1}{64} \sum_{k=1}^{64} P_k\end{aligned}$$

...(11)

where P_k is a pixel value in a luminance signal block of an original picture. The purpose of taking a minimum value in the equation (11) is to refine the quantization if there is a monotonous pattern portion even in a portion of the macroblock.

The normalized activity $Nact_j$, the value of which assumes a value in a range from 0.5 to 2, is found by the equation (12):

$$Nact_j = \frac{2 \times act_j + avg_act}{act_j + 2 \times avg_act} \quad \dots(12)$$

where avg_act is an average value of act_j in a picture encoded directly previously.

The quantization scale code $mquant_j$, which takes psychoacoustic characteristics into account, is given, based on the quantization scale code Q_j obtained at step S12, in accordance with the following equation (13):

$$mquant_j = Q_j \times Nact_j \quad \dots(13).$$

The above-described code volume controlling system, defined in MPEG2 Test Model 5, is known to suffer from the following limitations, such that, in actual control, measured_s need to be taken against these limitations. That is, the first limitation is that the first step S11 cannot cope with a scene change and that, after a scene change, the parameter avg_act used at step S13 takes on an incorrect value, after a scene change.

The second limitation is that there is no assurance that the constraint condition of VBV (video buffer verifier), as provided in MPEG2 and MPEG4, can be met.

Meanwhile, in the execution of the equation (11), it is necessary to calculate the totality of the average values and variance values of the pixel values for each macroblock, thus necessitating voluminous processing operations. There also are also occasions where the fact that avg_act in the equation (12) is not an average value in the frame but is an average value in the directly previous frame and it obstructs stable coderate control.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide a picture information converting method and apparatus whereby the processing volume in calculating the activity is diminished to assure stabilized coderate control.

The present invention provides a method and apparatus for converting ~~the~~ interlaced scanned compressed picture information, compressed in accordance with a first compression coding system, into ~~the~~ progressive scanned output compressed picture information, compressed in accordance with a second compression coding system. The second activity information of a pixel block constituting a frame of the output compressed picture information is synthesized with the use of the first activity information constituting a frame of the input compressed picture information. The second activity information, so synthesized, is used as a parameter of adaptive

quantization at the time of compression in the second compression encoding system.

According to the present invention, the interlaced scanned MPEG2 compressed picture information (bitstream) is the input compressed picture information, ~~whilst~~ while the progressive scanned MPEG4 compressed picture information (bitstream) is the output compressed picture information. Each of the MPEG2 compressed picture information and the MPEG4 compressed picture information is constituted by pixel blocks, that is macroblocks, each being composed of plural pixels.

According to the present invention, the interlaced scanned MPEG2 compressed picture information (bitstream) is an input. There are provided a picture type decision unit, an MPEG2 picture information decoding unit (I/P picture), a decimating unit, a delay buffer, an MPEG4 picture information encoding unit (I/P-VOP), a motion vector synthesis unit, a motion vector detection unit, an information buffer and an activity synthesis unit. Using the pixel block based, that is macroblock-based, activity information extracted from the MPEG2 compressed picture information (bitstream) as the input compressed picture information, the MPEG4 present invention encoding (I/P-VOP) is performed to output the MPEG4 compressed picture information (bitstream), serving as progressive scanned output compressed picture information, in an optimized macroblock-based coderate allocation, with a smaller processing volume. It also is ~~also~~ possible to eliminate the delay buffer 10 and ~~to~~ provide a compressed information analysis unit.

In the above-described structure, the picture type decision unit leaves only the I/P picture relevant portion in the input MPEG2 compressed picture information (bitstream), as the B-picture relevant portion is discarded. The compressed picture information (bitstream) pertinent to the I/P picture, output from the picture type decision unit, is decoded, using all of the order eight DCT coefficients, or only low frequency coefficients, for both the horizontal and vertical directions. The decimating unit takes out only the first field or the second field of the picture information output from the MPEG2 picture information decoding unit (I/P picture) for conversion to the progressive scanned picture, while performing downsampling for conversion to a desired picture frame size. The delay buffer stores the picture information for one frame. The MPEG4 picture information encoding unit (I/P-VOP) encodes the picture information output from the delay buffer in accordance with the MPEG4 encoding system. The motion vector synthesis unit effects mapping into motion vector values corresponding to the scanning-converted picture data based on the motion vector value in the input compressed picture information (bitstream) as detected by the MPEG2 picture information decoding unit (I/P picture). The motion vector detection unit detects the motion vector to high precision based on the motion vector value output from the motion vector synthesis unit. The information buffer extracts the macroblock-based activity information, obtained in performing decoding in the MPEG2 picture information decoding unit (I/P picture), and to stores the extracted

information therein. The activity synthesis unit synthesizes, from the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream), stored in the information buffer, the macroblock-based activity information in the output MPEG4 compressed picture information (bitstream) ~~to and~~ transmits the synthesized activity information to the MPEG4 picture information encoding unit (I/P-VOP).

According to the present invention, as described above, the interlaced scanned MPEG2 compressed picture information (bitstream) is used as input and, from the activity information for respective macroblocks, the activity information for each macroblock in the output MPEG4 compressed picture information (bitstream) is synthesized and used for adaptive quantization. The input MPEG2 compressed picture information (bitstream) may be converted in this manner into the progressive scanned MPEG4 compressed picture information (bitstream) in optimized macroblock-based coderate allocation, with a smaller processing volume.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig.1 is a block diagram showing the structure of a picture information converting apparatus according to a first embodiment of the present invention.

Fig.2 illustrates a method for generating the activity information Act_j .

Fig.3 is a flowchart for illustrating the operating principle of the quantization processing used.

Fig.4 is a block diagram showing the structure of a picture information converting apparatus according to a second embodiment of the present invention.

Fig.5 is a block diagram showing the structure of a conventional picture information converting apparatus.

Fig.6 is a flowchart showing the operating principle of a conventional encoding controlling system.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to the drawings, preferred embodiments of the present invention will be explained in detail.

First, a picture information converting apparatus according to a first embodiment of the present invention is explained.

Referring to Fig.3, the picture information converting apparatus includes a picture type decision unit 7, an MPEG2 picture information decoding unit (I/P picture) 8, a decimating unit 9, a delay buffer 10, an MPEG2 picture information encoding unit (I/P-VOP) 11, a motion vector synthesizing unit 12, an information buffer 14 and an activity synthesis unit 14.

This picture information converting apparatus is fed with the interlaced scanned MPEG2 compressed picture information (bitstream) made up of an intra-coded picture (I-picture) obtained on intra-frame coding, a forward predicted picture (P-picture) obtained on predictive coding by referring to a ~~the~~ forward direction in the display

sequence, and a bi-directionally coded picture (B-picture) obtained on predictive coding by referring to the forward and backward directions in the display sequence.

This MPEG2 compressed picture information (bitstream) is discriminated in the picture type decision unit 1 as to whether it is of an I/P picture or of a B-picture. Only the I/P picture is output to the next following MPEG2 picture information decoding unit (I/P picture) 2, while the B-picture is discarded.

The MPEG2 picture information decoding unit (I/P) 8 decodes the MPEG2 compressed picture information (bitstream) into picture signals, while extracting the activity information, and ~~to~~ routes the resulting information to the information buffer 14. Since the data on the B-picture has been discarded in the picture type decision unit 7, it is sufficient for the MPEG2 picture information decoding unit (I/P) 8 to have the function only of decoding the I/P picture.

The pixel value output by the MPEG2 picture information decoding unit (I/P picture) 8 is input to the decimating unit 9, which then decimates the pixel value by 1/2 in the horizontal direction, while leaving only one of the data of the first field and the data of the second field ~~to~~and discards the other. By this decimation, there is produced a sequentially scanned picture having a size equal to 1/4 of the input picture information.

Meanwhile, for encoding athe picture output from the decimating unit 9 in the MPEG4 picture information encoding unit (I/P-VOP) 11 in terms of a macro-block

composed of 16×16 pixels as a unit, the number of pixels of the picture needs to be multiples of 16 in both the horizontal and vertical directions. The decimating unit 9 interpolates or discards the pixels simultaneously with decimation.

For example, if the input MPEG2 compressed picture information (bitstream) conforms to the standard of the NTSC (National Television System Committee), that is an interlaced scanned picture with 720×480 pixels and 30Hz, the as-decimated picture frame is of the SIF size (360×240 pixels). This picture is turned into 352×240 pixels by discarding eight lines, e.g., at the right or left end in the horizontal direction in the decimating unit 9.

The picture may also be converted into, e.g., a picture of the QSIF size (176×112 pixels), which is a picture frame of approximately $1/4 \times 1/4$, by changing the operation in the decimating unit 9.

The above-mentioned reference 1 is directed to a picture information converting apparatus in which, the processing in the MPEG2 picture information decoding unit (I/P) 8 is the decoding operation employing all of the order eight DCT coefficients in the input MPEG2 compressed picture information for both the horizontal and vertical directions. The apparatus shown in Fig.1 is not limited to this configuration. For example, the apparatus shown in Fig.3 also may ~~also~~ be designed to execute decoding using only low-frequency components of the order eight DCT coefficients for only the horizontal or vertical direction of both the horizontal and vertical directions to

suppress the deterioration in the picture quality to a the minimum as well as and to decrease the processing volume accompanying the decoding processing and the video memory capacity.

The progressive scanned picture, generated by the decimating unit 9, is delayed one frame by the delay buffer 10 and subsequently encoded by the MPEG4 picture information encoding unit (I/P-VOP) 11 into an intra-frame-coded I-VOP and ~~into~~ a P-VOP, predictively coded by referring to the forward direction in the display sequence, so as to be output as the MPEG4 compressed picture information (bitstream).

The motion vector information in the input MPEG2 compressed picture information (bitstream) is mapped in the motion vector synthesizing unit 12 into the a motion vector with respect to the as-decimated picture information. A motion vector detection unit 13 detects the high-precision motion vector based on the motion vector value, synthesized in the motion vector synthesizing unit 12.

The VOP means a video object plane and is equivalent to a frame in MPEG2. The I-VOP, P-VOP and B-VOP means an intra-coded VOP corresponding to an I-picture, a forward predictive-coded VOP corresponding to the a P-picture and a bi-directionally predictive-coded VOP corresponding to the a B-picture, respectively.

In this picture information converting apparatus, the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) is sent from the MPEG2 picture information decoding unit (I/P) 8 to the information buffer 14

so as to be stored by one frame therein. The activity information as-used here may be one of those as-found by the following six methods:

The first method uses the macroblock-based quantization scale Q in the MPEG2 compressed picture information (bitstream). The second method is to use the code volume (number of bits) allocated to the macro-block based luminance component DCT coefficients in the input MPEG2 compressed picture information (bitstream). The third method is to use the code volume (number of bits) allocated to the macro-block based DCT coefficients in the input MPEG2 compressed picture information (bitstream). The fourth method is to use the code volume (number of bits) allocated to the macro-block in the input MPEG2 compressed picture information (bitstream). The fifth method is to use X as given by the following equation (14):

$$X = Q \cdot B \quad \dots(14)$$

where B is the code volume (number of bits) allocated to each macroblock in the input MPEG2 compressed picture information (bitstream).

It is noted that B may be the entire code volume (number of bits) allocated to a macroblock, the code volume (number of bits) allocated to the DCT coefficients or the code volume (number of bits) allocated to the luminance component DCT coefficients; and, Q is a quantization scale. The sixth method is to use non-zero DCT coefficients for luminance components or both the luminance and chrominance

components in each macroblock in the input MPEG2 compressed picture information (bitstream).

In the following, it is assumed that the MPEG4 compressed picture information (bitstream) of the progressive scanned picture ~~having~~has a 1/4 picture frame of the MPEG2 compressed picture information (bitstream) of the input progressive scanned picture.

The activity information Act_j for a given macroblock in the output MPEG4 compressed picture information (bitstream) shown at B in Fig.4 is generated from the activity information $Act_{j,n}$, where $n = 1, \dots, 4$, in the input MPEG2 compressed picture information (bitstream) shown at A in Fig.4, using an activity synthesis unit 15, in accordance with the following equation (15):

$$Act_j = (Act_{j,1}, Act_{j,2}, Act_{j,3}, Act_{j,4}) \dots (15)$$

where the function f may be such a one that outputs an average value, if an input sample, or a minimum value.

The activity synthesis unit 15 calculates the aforementioned macroblock-based information Act_j for the output MPEG4 compressed picture information (bitstream) and an average value Avg_act over the entire VOP of Act_j to output the result to the MPEG4 picture information encoding unit (I/P-VOP) 11. For calculating Avg_act , it is necessary to know Act_j over the entire picture displayed. To this end, the delay

buffer 10 is used.

The MPEG4 picture information encoding unit (I/P-VOP) 11 calculates the normalized activity $N Act_j$ for each macroblock_j using the parameter Act_j as calculated in the activity synthesis unit 15, and the Avg_act , as shown by the following equation (16), in association with the equation (12) to execute the macroblock-based adaptive quantization processing.

$$Nact_j = \frac{2 \times Act_j + Avg_act}{Act_j + 2 \times Avg_act} \quad \dots(16)$$

~~in association with the equation (12) to execute the macroblock-based adaptive quantization processing.~~

The sequence of processing operations up to this adaptive quantization is now explained by referring to Fig.5.

At the first step S21, the macroblock based activity information $Act_{j,n}$ in the input MPEG2 compressed picture information (bitstream); output from the MPEG2 picture information decoding unit (I/P) 8; is stored in the information buffer 14.

At step S22, the activity synthesis unit 15 generates the activity information Act_j for a macroblock in the MPEG4 compressed picture information (bitstream); from the activity information Act_j stored in the information buffer 14.

At step S23, the activity synthesis unit 15 calculates the average value Avg_act of the activity information Act_j . At step S24, the activity synthesis unit 15 calculates

the normalized activity $Nact_j$.

At step S25, the MPEG4 picture information encoding unit 11 executes present invention encoding, using the adaptive quantization, based on the normalized activity $Nact_j$ supplied from the activity synthesis unit 15.

By executing the above processing, the execution of the equation (11) becomes unnecessary, thus diminishing the processing volume. On the other hand, avg_act in the equation (11) is an average value for the directly previous VOP, while Avg_act in the equation (16) is an average value of the VOP in question, thus assuring more stabilized coderate control.

The picture information processing apparatus according to a second embodiment of the present invention is hereinafter explained.

Referring to Fig.6, the present picture information converting apparatus includes a picture type decision unit 16, a compressed information analysis unit 17, an MPEG2 picture information decoding unit (I/P picture) 18, a decimating unit 19, an MPEG4 picture information encoding unit (I/P-VOP) 20, a motion vector synthesis unit 21, a motion vector detection unit 22, an information buffer 23 and an activity synthesis unit 24. In the picture information converting apparatus of the first embodiment, shown in Fig.3, the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) is extracted in the MPEG2 picture information decoding unit (I/P) 8 and the one-frame delay is introduced in the delay

buffer 10₃; whereas, in the present picture information converting apparatus, the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) is extracted, at the same time as the one-frame delay is introduced, in the compressed information analysis unit 17.

Other features of the present embodiment ~~are~~ is the same as those of the first embodiment described above; and ~~is~~ are not explained for clarity.

In the above-described embodiment, the macroblock-based information in the input MPEG2 compressed picture information (bitstream), extracted in the MPEG2 picture information decoding unit (I/P picture), is used to reduce the processing volume to realize stabilized code rate control.

In the foregoing, the MPEG2 compressed picture information (bitstream) and the MPEG4 compressed picture information (bitstream) are used as the input and the output, respectively. However, the input or the output is not limited to these and may also be other types of ~~the~~ compressed picture information (bitstream), such as ~~the~~ MPEG-1 or H.263.

Stable code rate control is performed in converting the picture information. The MPEG2 interlaced scanned compressed picture information (bitstream) is to be converted into the progressive scanned MPG4 compressed picture information (bitstream). An activity synthesis unit 15 synthesizes, from the activity-based activity information in the MPEG2 compressed picture information (bitstream), the macroblock-based activity information in the MPEG4 compressed picture information (bitstream). An MPEG4 picture information encoding unit (I/P-VOP) 11 uses the so-synthesized activity information as the parameter for adaptive quantization at the time of encoding the MPEG4 picture information.



80001-2107
CLEAN VERSION

Attorney Docket No.: SON-2107
Application No.: 09/862,421

TITLE OF THE INVENTION

Picture Information Conversion Method and Apparatus

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to a method and apparatus for converting picture information. More particularly, it relates to a picture information conversion method and apparatus for use in receiving picture information of, e.g., MPEG pictures compressed by orthogonal transform, such as discrete cosine transform, and motion compression (bitstream), over a broadcast satellite, cable TV or a network medium, such as the Internet, or in processing the bitstream on a recording medium, such as an optical disc or a magneto-optical disc.

Description of Related Art

There has so far been presented a picture information compression system, such as MPEG, for compressing the picture information by motion compression, by exploiting the redundancy proper to the picture information, and with a view to handling the picture information as digital data and to high-efficiency transmission and storage of the information. The apparatus conforming to this picture information compression method is finding widespread use in information distribution by, e.g., a broadcasting station and in information reception in homes.

In particular, the MPEG2 (ISO/IEC 13818-2) is defined as being a

comprehensive picture encoding system applicable to both interlaced and progressive scanned pictures, to standard definition pictures, and to high-definition pictures.

That is, in the MPEG2 encoding compression system, codes of a bitrate of 4 to 8 Mbps are allocated to an interlaced scanned picture of a standard resolution with 720×480 pixels, and codes of a bitrate of 18 to 22 Mbps are allocated to a progressive scanned picture of a standard resolution with 1920×1088 pixels to realize a high compression factor and a high picture quality.

In light of the above, the MPEG2 is estimated to continue to be in extensive use in professional and consumer use.

However, the MPEG2 is mainly intended for high picture quality encoding for broadcasting, while it is not adapted to a coderate lower than that in MPEG1, that is, it is not adapted to an encoding system with a higher compression rate.

On the other hand, it may be predicted that the need for an encoding system with a higher compression rate will continue to increase. In order to cope with this situation, standardization of the MPEG4 encoding system with a high compression rate is underway. For this picture encoding system, the international standardization was acknowledged in December 1998 as ISO/IEC 14496-2.

Meanwhile, there also exists the need for converting the MPEG2 compressed picture information, once encoded for digital broadcast, into compressed picture information (bitstream) of a lower code rate more amenable to processing on a

For accommodating these needs, there is presented a picture information converting apparatus (transcoder) in "Field-to-Frame Transcoding with Spatial and Temporal Downsampling" (Susie L. Wee, John G. Apostolopoulos, and Nick Feamster, ICIP 99; referred to below as reference 1).

As shown in Fig.1, the picture information converting apparatus (transcoder) presented in this reference 1 is made up of a picture type decision unit 1, an MPEG2 picture information decoding unit (I/P picture) 2, a decimating unit 3, an MPEG2 picture information encoding unit (I/P-VOP) 4, a motion vector synthesis unit 5 and a motion vector detection unit 6.

This picture information converting apparatus is fed with the interlaced scanned MPEG2 compressed picture information (bitstream) made up of an intra-coded picture (I-picture) obtained on intra-frame coding, a forward predicted picture (P-picture) obtained on predictive coding by referring to a forward direction in the display sequence, and a bi-directionally coded picture (B-picture) obtained on predictive coding by referring to the forward and backward directions in the display sequence.

This MPEG2 compressed picture information (bitstream) is discriminated in the picture type decision unit 1 as to whether it is of an I/P picture or of a B-picture. Only the I/P picture is output to the next following MPEG2 picture information decoding unit (I/P picture) 2, while the B-picture is discarded.

Similarly to the processing in a routine MPEG2 picture information decoding unit, the processing in the MPEG2 picture information decoding unit (I/P picture) 2 decodes the MPEG2 compressed picture information (bitstream) into picture signals,

The pixel value output by the MPEG2 picture information decoding unit (I/P picture) 2 is input to the decimating unit 3, which then decimates the pixel value by $1/2$ in the horizontal direction, while leaving only one of the data of the first field and the data of the second field and discarding the other. By this decimation, there is produced a progressive scanned picture having a size equal to $1/4$ of the input picture information.

The progressive scanned picture generated by the decimating unit 3 is encoded by the MPEG2 picture information encoding unit (I/P-VOP) 4 into an intra-frame-coded I-VOP and a P-VOP obtained on predictive coding by referring to the forward direction in the display sequence, and it is output as the MPEG4 compressed picture information (bitstream). Meanwhile, VOP means a video object plane and is equivalent to a frame in MPEG2.

The motion vector information in the input MPEG2 compressed picture information (bitstream) is mapped in the motion vector synthesis unit 5 into a motion vector for the as-decimated picture information. The motion vector detection unit 6 detects the high precision motion vector based on the motion vector value synthesized

in the motion vector synthesis unit 5.

\Reference 1 discusses a picture information converting apparatus for generating the MPEG2 compressed picture information (bitstream) having a size equal to $1/2 \times 1/2$ of the input MPEG2 compressed picture information (bitstream). That is, if the input MPEG2 compressed picture information (bitstream) is in meeting with the NTSC (National Television System Committee), the output MPEG4 compressed picture information (bitstream) is of an SIF size (352×240 pixels).

Meanwhile, in the picture information converting apparatus shown in Fig.1, the code rate control in the MPEG4 picture information encoding unit (I/P-VOP) 4 represents a significant factor in determining the picture quality in the MPEG4 compressed picture information. In the ISO/IEC 14496-2, there is no particular definition as to the coderate controlling system, such that each vendor may use a system that is possibly optimal from the viewpoint of the processing volume and the output picture quality depending on the particular application. The system discussed in MPEG2 Test Model 5(ISO/IEC JTC1/SC29/WG11 NO400) is hereinafter explained as a typical coderate controlling system.

The code rate control flow is now explained by referring to the flowchart of Fig.2. At a first step S11, the picture information encoding unit (I/P-VOP) 4 allocates bits to each picture, with the target code rate (target bitrate) and the GOP (group of pictures) as input variables. It is noted that a GOP means a set of pictures accessible at

That is, at step S11, the picture information encoding unit (I/P-VOP) 4 distributes bits to be allocated to each picture in the GOP based on the volume of bits allocated to a picture not as yet decoded in the GOP inclusive of pictures intended for allocation. This bit volume is referred to below as R. This distribution is repeated in the sequence of the encoded pictures in the GOP. In this case, coderate allocation to each picture is made using two assumptions as now explained.

It is first assumed that the product of an average quantization scale code used in encoding each picture and the volume of the codes generated is unchanged from one picture type to another as long as the picture displayed is not changed. Based on this supposition, variables X_i , X_p and X_b representing the picture complexity (global complexity measure) are updated after encoding each picture in accordance with the following equation (1) from one picture type to another:

$$X_i = S_i \cdot Q_i$$

$$X_p = S_p \cdot Q_p$$

$$X_b = S_b \cdot Q_b$$

...(1)

It is noted that S_i , S_p and S_b denote the volumes of the codes generated on picture encoding, and Q_i , Q_p and Q_b are average quantization scale codes at the time of

picture encoding. On the other hand, the initial value, in terms of the target bitrate bit_rate [bits/sec], is as indicated in the following equation (2):

$$X_i = 160 \times \text{bit_rate} / 115$$

$$X_p = 60 \times \text{bit_rate} / 115$$

$$X_b = 42 \times \text{bit_rate} / 115$$

...(2).

Second, it is assumed that the overall picture quality is optimized at all times when the proportions K_p , K_b of the quantization scale code of the P- and B-pictures, referenced to the quantization scale code of an I-picture, are of values defined in the following equation (3):

$$K_p = 1.0; K_b = 1.4$$

...(3).

That is, the quantization scale code of a B-picture is set at all times so as to be 1.4 times the quantization scale codes of the I- and P-pictures. This is, based on the supposition that if the volume of the codes that can be saved in a B-picture by encoding the B-picture slightly more coarsely than the I- and P-pictures is added to the code volume of the I- and P-pictures, the I- and P-pictures can be improved in picture quality, so that the B-picture which refers to these also can be improved in picture quality.

From the above-mentioned two assumptions, the volumes of bits allocated to

each picture of the GOP (T_i , T_p , T_b) are as indicated by the following equation (4):

$$T_i = \max \left\{ \frac{R}{1 + \frac{N_p \cdot X_p}{X_i \cdot K_p} + \frac{N_b \cdot X_b}{X_i \cdot K_b}}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

$$T_p = \max \left\{ \frac{R}{N_p + \frac{N_b \cdot K_p \cdot X_b}{K_b \cdot X_p}}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

$$T_b = \max \left\{ \frac{R}{N_b + \frac{N_p \cdot K_b \cdot X_p}{K_p \cdot X_b}}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

...(4).

where N_p and N_b denote the number of P- and B-pictures, respectively, not as yet encoded in the GOP.

Based on the value of the allocated codes thus found, the volume of bits R allocated to uncoded pictures in a GOP is updated in accordance with the following equation (5) each time a picture is encoded in accordance with steps 511 and 512:

$$R = R - S_{i,p,b}$$

...(5)

On the other hand, in encoding the first picture of the GOP, R is updated in accordance with the equation (6):

$$R = \frac{bit_rate \times N}{picture_rates} + R \quad \dots(6)$$

where N is the number of pictures in a GOP. The initial value of R at the beginning of a sequence is 0.

Then, at step S12, the picture information encoding unit (I/P-VOP) 4 performs rate control using a virtual buffer. That is, at step S12, the picture information encoding unit (I/P-VOP) 4 finds the quantization scale code by macro-block based feedback control, based on the capacitance of three types of the virtual buffer set independently for the respective pictures, in order to make the volume of allocated bits for the respective pictures found by the equation (4) at step S11 (T_i , T_p , T_b) coincident with the actual volume of generated codes.

Before proceeding to the encoding of the jth macroblock, the occupancy volume of the virtual buffer is found by the following equation (7):

$$\begin{aligned} d_j^i &= d_o^i + B_{j-1} - \frac{T_i \times (j-1)}{MB_cnt} \\ d_j^p &= d_o^p + B_{j-1} - \frac{T_p \times (j-1)}{MB_cnt} \\ d_j^b &= d_o^b + B_{j-1} - \frac{T_b \times (j-1)}{MB_cnt} \end{aligned}$$

... (7)

It is noted that do^i , do^p , do^b are initial occupancy volumes of the virtual buffers, B_j is the volume of bits generated from the leading end of a picture up to the j th macroblock and MB_cnt is the number of macroblocks in one picture. The occupancy of the virtual buffer at the time of end of encoding of each picture (dMB_cnt^i , dMB_cnt^p , dMB_cnt^b) is used as initial values (do^i , do^p , do^b) of the occupancy of the virtual buffer for the next picture in the same picture type.

The quantization scale code for the j th macroblock is then calculated in accordance with equation (8):

$$Q_j = \frac{d_j \times 31}{r} \quad \dots(8)$$

where r is a variable controlling the feedback loop response, termed a reaction parameter, and is given by equation (9):

$$r = 2 \times \frac{bit_rate}{picture_rate} \quad \dots(9).$$

Meanwhile, the initial value of the virtual buffer at the time of beginning the encoding is given by equation (10):

$$\begin{aligned}d_o^i &= 10 \times \frac{r}{31} \\d_o^p &= K_p \bullet d_o^i \\d_o^b &= K_{pb} \bullet d_o^i\end{aligned}\dots(10),$$

Finally, at S 13, the picture information encoding unit (I/P-VOP) 4 performs macro-block based adaptive quantization taking psychoacoustic characteristics into account. That is, at step S13, the picture information encoding unit (I/P-VOP) 4 varies the quantization scale code found at step S12 by a variable termed macroblock-based activity in such a manner that the quantization scale code will be quantized finely and coarsely in a monotonous pattern portion where deterioration tends to be visually outstanding and in a complex pattern portion where deterioration is less likely to be outstanding, respectively.

The activity is given, using luminance signal pixel values of an original picture, four blocks in the frame DCT mode and four blocks in the field DCT mode, totaling eight blocks, by the following equation (11):

$$\begin{aligned}act_j &= 1 + \min_{sblk=1,8} (var_sblk) \\var_sblk &= \frac{1}{64} \sum_{k=1}^{64} (P_k - \bar{P})^2 \\ \bar{P} &= \frac{1}{64} \sum_{k=1}^{64} P_k\end{aligned}$$

...(11)

where P_k is a pixel value in a luminance signal block of an original picture. The purpose of taking a minimum value in equation (11) is to refine the quantization if there is a monotonous pattern portion even in a portion of the macroblock.

The normalized activity $Nact_j$, the value of which assumes a value in a range from 0.5 to 2, is found by equation (12):

$$Nact_j = \frac{2 \times act_j + avg_act}{act_j + 2 \times avg_act} \quad \dots(12)$$

where avg_act is an average value of act_j in a picture encoded directly previously.

The quantization scale code $mquant_j$, which takes psychoacoustic characteristics into account, is given based on the quantization scale code Q_j obtained at step S12 in accordance with the following equation (13):

$$mquant_j = Q_j \times Nact_j \quad \dots(13).$$

The above-described code volume controlling system, defined in MPEG2 Test Model 5, is known to suffer from the following limitations, such that, in actual control, measures need to be taken against these limitations. That is, the first limitation is that the first step S11 cannot cope with a scene change and, after a scene change, the parameter avg_act used at step S13 takes on an incorrect value. The second limitation

is that there is no assurance that the constraint condition of VBV (video buffer verifier), as provided in MPEG2 and MPEG4, can be met.

Meanwhile, in the execution of the equation (11), it is necessary to calculate the totality of the average values and variance values of the pixel values for each macroblock, thus necessitating voluminous processing operations. There also are occasions where the fact that avg_act in the equation (12) is not an average value in the frame but is an average value in the directly previous frame and it obstructs stable coderate control.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide a picture information converting method and apparatus whereby the processing volume in calculating the activity is diminished to assure stabilized coderate control.

The present invention provides a method and apparatus for converting interlaced scanned compressed picture information, compressed in accordance with a first compression coding system, into progressive scanned output compressed picture information, compressed in accordance with a second compression coding system. The second activity information of a pixel block constituting a frame of the output compressed picture information is synthesized with the use of the first activity information constituting a frame of the input compressed picture information. The second activity information, so synthesized, is used as a parameter of adaptive

quantization at the time of compression in the second compression encoding system.

According to the present invention, the interlaced scanned MPEG2 compressed picture information (bitstream) is the input compressed picture information, while the progressive scanned MPEG4 compressed picture information (bitstream) is the output compressed picture information. Each of the MPEG2 compressed picture information and the MPEG4 compressed picture information is constituted by pixel blocks, that is macroblocks, each being composed of plural pixels.

According to the present invention, the interlaced scanned MPEG2 compressed picture information (bitstream) is an input. There are provided a picture type decision unit, an MPEG2 picture information decoding unit (I/P picture), a decimating unit, a delay buffer, an MPEG4 picture information encoding unit (I/P-VOP), a motion vector synthesis unit, a motion vector detection unit, an information buffer and an activity synthesis unit. Using the pixel block based, that is macroblock-based, activity information extracted from the MPEG2 compressed picture information (bitstream) as the input compressed picture information, the MPEG4 present invention encoding (I/P-VOP) is performed to output the MPEG4 compressed picture information (bitstream), serving as progressive scanned output compressed picture information, in an optimized macroblock-based coderate allocation with a smaller processing volume. It also is possible to eliminate the delay buffer 10 and provide a compressed information analysis unit.

In the above-described structure, the picture type decision unit leaves only the I/P picture relevant portion in the input MPEG2 compressed picture information (bitstream), as the B-picture relevant portion is discarded. The compressed picture information (bitstream) pertinent to the I/P picture output from the picture type decision unit is decoded using all of the order eight DCT coefficients, or only low frequency coefficients, for both the horizontal and vertical directions. The decimating unit takes out only the first field or the second field of the picture information output from the MPEG2 picture information decoding unit (I/P picture) for conversion to the progressive scanned picture while performing downsampling for conversion to a desired picture frame size. The delay buffer stores the picture information for one frame. The MPEG4 picture information encoding unit (I/P-VOP) encodes the picture information output from the delay buffer in accordance with the MPEG4 encoding system. The motion vector synthesis unit effects mapping into motion vector values corresponding to the scanning-converted picture data based on the motion vector value in the input compressed picture information (bitstream) detected by the MPEG2 picture information decoding unit (I/P picture). The motion vector detection unit detects the motion vector to high precision based on the motion vector value output from the motion vector synthesis unit. The information buffer extracts the macroblock-based activity information obtained in performing decoding in the MPEG2 picture information decoding unit (I/P picture) and stores the extracted

information therein. The activity synthesis unit synthesizes from the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) stored in the information buffer the macroblock-based activity information in the output MPEG4 compressed picture information (bitstream) and transmits the synthesized activity information to the MPEG4 picture information encoding unit (I/P-VOP).

According to the present invention, as described above, the interlaced scanned MPEG2 compressed picture information (bitstream) is used as input and, from the activity information for respective macroblocks, the activity information for each macroblock in the output MPEG4 compressed picture information (bitstream) is synthesized and used for adaptive quantization. The input MPEG2 compressed picture information (bitstream) may be converted in this manner into the progressive scanned MPEG4 compressed picture information (bitstream) in optimized macroblock-based coderate allocation with a smaller processing volume.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig.1 is a block diagram showing the structure of a picture information converting apparatus according to a first embodiment of the present invention.

Fig.2 illustrates a method for generating the activity information Act_i .

Fig.3 is a flowchart for illustrating the operating principle of the quantization processing used.

Fig.4 is a block diagram showing the structure of a picture information converting apparatus according to a second embodiment of the present invention.

Fig.5 is a block diagram showing the structure of a conventional picture information converting apparatus.

Fig.6 is a flowchart showing the operating principle of a conventional encoding controlling system.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to the drawings, preferred embodiments of the present invention will be explained in detail.

First, a picture information converting apparatus according to a first embodiment of the present invention is explained.

Referring to Fig.3, the picture information converting apparatus includes a picture type decision unit 7, an MPEG2 picture information decoding unit (I/P picture) 8, a decimating unit 9, a delay buffer 10, an MPEG2 picture information encoding unit (I/P-VOP) 11, a motion vector synthesizing unit 12, an information buffer 14 and an activity synthesis unit 14.

This picture information converting apparatus is fed with the interlaced scanned MPEG2 compressed picture information (bitstream) made up of an intra-coded picture (I-picture) obtained on intra-frame coding, a forward predicted picture (P-picture) obtained on predictive coding by referring to the forward direction in the display

sequence, and a bi-directionally coded picture (B-picture) obtained on predictive coding by referring to the forward and backward directions in the display sequence.

This MPEG2 compressed picture information (bitstream) is discriminated in the picture type decision unit 1 as to whether it is of an I/P picture or of a B-picture. Only the I/P picture is output to the next following MPEG2 picture information decoding unit (I/P picture) 2, while the B-picture is discarded.

The MPEG2 picture information decoding unit (I/P) 8 decodes the MPEG2 compressed picture information (bitstream) into picture signals, while extracting the activity information, and routes the resulting information to the information buffer 14.

Since the data on the B-picture has been discarded in the picture type decision unit 7, it is sufficient for the MPEG2 picture information decoding unit (I/P) 8 to have the function only of decoding the I/P picture.

The pixel value output by the MPEG2 picture information decoding unit (I/P picture) 8 is input to the decimating unit 9, which then decimates the pixel value by 1/2 in the horizontal direction, while leaving only one of the data of the first field and the data of the second field and discards the other. By this decimation, there is produced a sequentially scanned picture having a size equal to 1/4 of the input picture information.

Meanwhile, for encoding the picture output from the decimating unit 9 in the MPEG4 picture information encoding unit (I/P-VOP) 11 in terms of a macro-block

composed of 16×16 pixels as a unit, the number of pixels of the picture needs to be multiples of 16 in both the horizontal and vertical directions. The decimating unit 9 interpolates or discards the pixels simultaneously with decimation.

For example, if the input MPEG2 compressed picture information (bitstream) conforms to the standard of the NTSC (National Television System Committee), that is an interlaced scanned picture with 720×480 pixels and 30Hz, the as-decimated picture frame is of the SIF size (360×240 pixels). This picture is turned into 352×240 pixels by discarding eight lines, e.g., at the right or left end in the horizontal direction in the decimating unit 9.

The picture may also be converted into, e.g., a picture of the QSIF size (176×112 pixels), which is a picture frame of approximately $1/4 \times 1/4$, by changing the operation in the decimating unit 9.

The above-mentioned reference 1 is directed to a picture information converting apparatus in which the processing in the MPEG2 picture information decoding unit (I/P) 8 is the decoding operation employing all of the order eight DCT coefficients in the input MPEG2 compressed picture information for both the horizontal and vertical directions. The apparatus shown in Fig.1 is not limited to this configuration. For example, the apparatus shown in Fig.3 also may be designed to execute decoding using only low-frequency components of the order eight DCT coefficients for only the horizontal or vertical direction of both the horizontal and vertical directions to

suppress the deterioration in the picture quality to a minimum and to decrease the processing volume accompanying the decoding processing and the video memory capacity.

The progressive scanned picture generated by the decimating unit 9 is delayed one frame by the delay buffer 10 and subsequently encoded by the MPEG4 picture information encoding unit (I/P-VOP) 11 into an intra-frame-coded I-VOP and a P-VOP, predictively coded by referring to the forward direction in the display sequence, so as to be output as the MPEG4 compressed picture information (bitstream).

The motion vector information in the input MPEG2 compressed picture information (bitstream) is mapped in the motion vector synthesizing unit 12 into a motion vector with respect to the as-decimated picture information. A motion vector detection unit 13 detects the high-precision motion vector based on the motion vector value synthesized in the motion vector synthesizing unit 12.

The VOP means a video object plane and is equivalent to a frame in MPEG2. The I-VOP, P-VOP and B-VOP mean an intra-coded VOP corresponding to an I-picture, a forward predictive-coded VOP corresponding to a P-picture and a bi-directionally predictive-coded VOP corresponding to a B-picture, respectively.

In this picture information converting apparatus, the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) is sent from the MPEG2 picture information decoding unit (I/P) 8 to the information buffer 14

so as to be stored by one frame therein. The activity information used here may be one of those found by the following six methods:

The first method uses the macroblock-based quantization scale Q in the MPEG2 compressed picture information (bitstream). The second method is to use the code volume (number of bits) allocated to the macro-block based luminance component DCT coefficients in the input MPEG2 compressed picture information (bitstream). The third method is to use the code volume (number of bits) allocated to the macro-block based DCT coefficients in the input MPEG2 compressed picture information (bitstream). The fourth method is to use the code volume (number of bits) allocated to the macro-block in the input MPEG2 compressed picture information (bitstream). The fifth method is to use X as given by the following equation (14):

$$X = Q \cdot B \quad \dots(14)$$

where B is the code volume (number of bits) allocated to each macroblock in the input MPEG2 compressed picture information (bitstream).

It is noted that B may be the entire code volume (number of bits) allocated to a macroblock, the code volume (number of bits) allocated to the DCT coefficients or the code volume (number of bits) allocated to the luminance component DCT coefficients; and, Q is a quantization scale. The sixth method is to use non-zero DCT coefficients for luminance components or both the luminance and chrominance components in

each macroblock in the input MPEG2 compressed picture information (bitstream).

In the following, it is assumed that the MPEG4 compressed picture information (bitstream) of the progressive scanned picture has a 1/4 picture frame of the MPEG2 compressed picture information (bitstream) of the input progressive scanned picture.

The activity information Act_j for a given macroblock in the output MPEG4 compressed picture information (bitstream) shown at B in Fig.4 is generated from the activity information $Act_{j,n}$, where $n = 1, \dots, 4$, in the input MPEG2 compressed picture information (bitstream) shown at A in Fig.4, using an activity synthesis unit 15 in accordance with the following equation (15):

$$Act_j = (Act_{j,1}, Act_{j,2}, Act_{j,3}, Act_{j,4}) \quad \dots(15)$$

where the function f may be one that outputs an average value, if an input sample, or a minimum value.

The activity synthesis unit 15 calculates the aforementioned macroblock-based information Act_j for the output MPEG4 compressed picture information (bitstream) and an average value Avg_act over the entire VOP of Act_j to output the result to the MPEG4 picture information encoding unit (I/P-VOP) 11. For calculating Avg_act , it is necessary to know Act_j over the entire picture displayed. To this end, the delay buffer 10 is used.

The MPEG4 picture information encoding unit (I/P-VOP) 11 calculates the

normalized activity $NAct_j$ for each macroblock using the parameter Act_j as calculated in the activity synthesis unit 15, and the Avg_act , as shown by the following equation (16), in association with the equation (12) to execute the macroblock-based adaptive quantization processing.

$$Nact_j = \frac{2 \times Act_j + Avg_act}{Act_j + 2 \times Avg_act} \quad \dots(16)$$

The sequence of processing operations up to this adaptive quantization is now explained by referring to Fig.5.

At the first step S21, the macroblock based activity information $Act_{j,n}$ in the input MPEG2 compressed picture information (bitstream) output from the MPEG2 picture information decoding unit (I/P) 8 is stored in the information buffer 14.

At step S22, the activity synthesis unit 15 generates the activity information Act_j for a macroblock in the MPEG4 compressed picture information (bitstream) from the activity information Act_j stored in the information buffer 14.

At step S23, the activity synthesis unit 15 calculates the average value Avg_act of the activity information Act_j . At step S24, the activity synthesis unit 15 calculates the normalized activity $Nact_j$.

At step S25, the MPEG4 picture information encoding unit 11 executes present invention encoding using the adaptive quantization based on the normalized activity

Nact_j supplied from the activity synthesis unit 15.

By executing the above processing, the execution of equation (11) becomes unnecessary, thus diminishing the processing volume. On the other hand, avg_act in equation (11) is an average value for the directly previous VOP, while Avg_act in equation (16) is an average value of the VOP in question, thus assuring more stabilized coderate control.

The picture information processing apparatus according to a second embodiment of the present invention is hereinafter explained.

Referring to Fig.6, the present picture information converting apparatus includes a picture type decision unit 16, a compressed information analysis unit 17, MPEG2 picture information decoding unit (I/P picture) 18, a decimating unit 19, MPEG4 picture information encoding unit (I/P-VOP) 20, a motion vector synthesis unit 21, a motion vector detection unit 22, an information buffer 23 and an activity synthesis unit 24. In the picture information converting apparatus of the first embodiment, shown in Fig.3, the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) is extracted in the MPEG2 picture information decoding unit (I/P) 8 and the one-frame delay is introduced in the delay buffer 10; whereas, in the present picture information converting apparatus, the macroblock-based activity information in the input MPEG2 compressed picture information (bitstream) is extracted, at the same time as the one-frame delay is

introduced, in the compressed information analysis unit 17.

Other features of the present embodiment are the same as those of the first embodiment described above and are not explained for clarity.

In the above-described embodiment, the macroblock-based information in the input MPEG2 compressed picture information (bitstream) extracted in the MPEG2 picture information decoding unit (I/P picture) is used to reduce the processing volume to realize stabilized code rate control.

In the foregoing, the MPEG2 compressed picture information (bitstream) and the MPEG4 compressed picture information (bitstream) are used as the input and the output, respectively. However, the input or the output is not limited to these and may also be other types of compressed picture information (bitstream), such as MPEG-1 or H.263.

Stable code rate control is performed in converting picture information. MPEG2 interlaced scanned compressed picture information (bitstream) is converted into the progressive scanned MPG4 compressed picture information (bitstream). An activity synthesis unit 15 synthesizes from the activity-based activity information in the MPEG2 compressed picture information (bitstream) the macroblock-based activity information in the MPEG4 compressed picture information (bitstream). An MPEG4 picture information encoding unit (I/P-VOP) 11 uses the so-synthesized activity information as the parameter for adaptive quantization at the time of encoding the MPEG4 picture information.